Why Machine Translation?

**Assimilation** — reader initiates translation, wants to know content

- user is tolerant of inferior quality
- focus of majority of research (GALE program, etc.)
Why Machine Translation?

Assimilation — reader initiates translation, wants to know content

- user is tolerant of inferior quality
- focus of majority of research (GALE program, etc.)

Communication — participants don’t speak same language, rely on translation

- users can ask questions, when something is unclear
- chat room translations, hand-held devices
- often combined with speech recognition, IWSLT campaign
Why Machine Translation?

**Assimilation** — reader initiates translation, wants to know content

- user is tolerant of inferior quality
- focus of majority of research (GALE program, etc.)

**Communication** — participants don’t speak same language, rely on translation

- users can ask questions, when something is unclear
- chat room translations, hand-held devices
- often combined with speech recognition, IWSLT campaign

**Dissemination** — publisher wants to make content available in other languages

- high demands for quality
- currently almost exclusively done by human translators
Why Machine Translation?

Assimilation — reader initiates translation, wants to know content

• user is tolerant of inferior quality
• focus of majority of research (GALE program, etc.)

Communication — participants don’t speak same language, rely on translation

• users can ask questions, when something is unclear
• chat room translations, hand-held devices
• often combined with speech recognition, IWSLT campaign

Dissemination — publisher wants to make content available in other languages

• high demands for quality
• currently almost exclusively done by human translators

OUR FOCUS
Goal: Helping Human Translators

If you can’t beat them, join them.
Goal: Helping Human Translators

If you can’t beat them, join them.

• How can machine translation help human translators?
Goal: Helping Human Translators

If you can’t beat them, join them.

• How can machine translation help human translators?

• First question: What do translators do?
Setup

- 10 students at the University of Edinburgh
  - half native French speakers
  - half native English speakers with advanced French
Setup

• 10 students at the University of Edinburgh
  – half native French speakers
  – half native English speakers with advanced French

• Each student translated
  – news stories
  – French-English
  – about 40 sentences
  – easy task: familiar content, no specialized terminology

• Keystroke log
Keystroke Log

Input:  Au premier semestre, l’avionneur a livr 97 avions.
Output: The manufacturer has delivered 97 planes during the first half.

(37.5 sec, 3.4 sec/word)

black: keystroke, purple: deletion, grey: cursor move
height: length of sentence
Analysis

- We can observe
  - slow typing
Analysis

- We can observe
  - slow typing
  - fast typing
Analysis

• We can observe
  – slow typing
  – fast typing
  – pauses
Analysis

• We can observe
  – slow typing
  – fast typing
  – pauses

• Pauses
  – beginning pause: reading the input sentence
  – final pause: reviewing the translation
Analysis

• We can observe
  – slow typing
  – fast typing
  – pauses

• Pauses
  – beginning pause: reading the input sentence
  – final pause: reviewing the translation
  – short pauses (2-6 seconds): hesitation
  – medium pauses (6-60 seconds): problem solving
  – big pauses (>60 seconds): serious problem
## Time Spent on Activities

<table>
<thead>
<tr>
<th>User</th>
<th>total</th>
<th>initial</th>
<th>final</th>
<th>short</th>
<th>medium</th>
<th>big</th>
<th>keystroke</th>
</tr>
</thead>
<tbody>
<tr>
<td>L2a</td>
<td>3.3s</td>
<td>0.1s</td>
<td>0.1s</td>
<td>0.2s</td>
<td>1.0s</td>
<td>0.1s</td>
<td>1.8s</td>
</tr>
<tr>
<td>L2b</td>
<td>7.7s</td>
<td>1.3s</td>
<td>0.1s</td>
<td>0.3s</td>
<td>1.8s</td>
<td>1.9s</td>
<td>2.3s</td>
</tr>
<tr>
<td>L2c</td>
<td>3.9s</td>
<td>0.2s</td>
<td>0.2s</td>
<td>0.3s</td>
<td>0.7s</td>
<td>-</td>
<td>2.5s</td>
</tr>
<tr>
<td>L2d</td>
<td>2.8s</td>
<td>0.2s</td>
<td>0.0s</td>
<td>0.2s</td>
<td>0.4s</td>
<td>0.1s</td>
<td>1.8s</td>
</tr>
<tr>
<td>L2e</td>
<td>5.2s</td>
<td>0.3s</td>
<td>0.0s</td>
<td>0.3s</td>
<td>1.9s</td>
<td>0.5s</td>
<td>2.2s</td>
</tr>
<tr>
<td>L1a</td>
<td>5.7s</td>
<td>0.5s</td>
<td>0.1s</td>
<td>0.3s</td>
<td>1.8s</td>
<td>0.7s</td>
<td>2.2s</td>
</tr>
<tr>
<td>L1b</td>
<td>3.2s</td>
<td>0.1s</td>
<td>0.1s</td>
<td>0.2s</td>
<td>0.4s</td>
<td>0.1s</td>
<td>2.2s</td>
</tr>
<tr>
<td>L1c</td>
<td>5.8s</td>
<td>0.3s</td>
<td>0.2s</td>
<td>0.5s</td>
<td>1.5s</td>
<td>0.3s</td>
<td>3.1s</td>
</tr>
<tr>
<td>L1d</td>
<td>3.4s</td>
<td>0.7s</td>
<td>0.1s</td>
<td>0.3s</td>
<td>0.6s</td>
<td>-</td>
<td>1.8s</td>
</tr>
<tr>
<td>L1e</td>
<td>2.8s</td>
<td>0.3s</td>
<td>0.2s</td>
<td>0.2s</td>
<td>0.3s</td>
<td>0.1s</td>
<td>1.9s</td>
</tr>
</tbody>
</table>

L2 = native French, L1 = native English
average time per input word
## Time Spent on Activities

<table>
<thead>
<tr>
<th>User</th>
<th>total</th>
<th>initial</th>
<th>final</th>
<th>short</th>
<th>medium</th>
<th>big</th>
<th>keystroke</th>
</tr>
</thead>
<tbody>
<tr>
<td>L2a</td>
<td>3.3s</td>
<td>0.1s</td>
<td>0.1s</td>
<td>0.2s</td>
<td>1.0s</td>
<td>0.1s</td>
<td>1.8s</td>
</tr>
<tr>
<td>L2b</td>
<td>7.7s</td>
<td>1.3s</td>
<td>0.1s</td>
<td>0.3s</td>
<td>1.8s</td>
<td>1.9s</td>
<td>2.3s</td>
</tr>
<tr>
<td>L2c</td>
<td>3.9s</td>
<td>0.2s</td>
<td>0.2s</td>
<td>0.3s</td>
<td>0.7s</td>
<td>-</td>
<td>2.5s</td>
</tr>
<tr>
<td>L2d</td>
<td>2.8s</td>
<td>0.2s</td>
<td>0.0s</td>
<td>0.2s</td>
<td>0.4s</td>
<td>0.1s</td>
<td>1.8s</td>
</tr>
<tr>
<td>L2e</td>
<td>5.2s</td>
<td>0.3s</td>
<td>0.0s</td>
<td>0.3s</td>
<td>1.9s</td>
<td>0.5s</td>
<td>2.2s</td>
</tr>
<tr>
<td>L1a</td>
<td>5.7s</td>
<td>0.5s</td>
<td>0.1s</td>
<td>0.3s</td>
<td>1.8s</td>
<td>0.7s</td>
<td>2.2s</td>
</tr>
<tr>
<td>L1b</td>
<td>3.2s</td>
<td>0.1s</td>
<td>0.1s</td>
<td>0.2s</td>
<td>0.4s</td>
<td>0.1s</td>
<td>2.2s</td>
</tr>
<tr>
<td>L1c</td>
<td>5.8s</td>
<td>0.3s</td>
<td>0.2s</td>
<td>0.5s</td>
<td>1.5s</td>
<td>0.3s</td>
<td>3.1s</td>
</tr>
<tr>
<td>L1d</td>
<td>3.4s</td>
<td>0.7s</td>
<td>0.1s</td>
<td>0.3s</td>
<td>0.6s</td>
<td>-</td>
<td>1.8s</td>
</tr>
<tr>
<td>L1e</td>
<td>2.8s</td>
<td>0.3s</td>
<td>0.2s</td>
<td>0.2s</td>
<td>0.3s</td>
<td>0.1s</td>
<td>1.9s</td>
</tr>
</tbody>
</table>

L2 = native French, L1 = native English
average time per input word

---

**Not much time**
### Time Spent on Activities

<table>
<thead>
<tr>
<th>User</th>
<th>total</th>
<th>initial</th>
<th>final</th>
<th>short</th>
<th>medium</th>
<th>big</th>
<th>keystroke</th>
</tr>
</thead>
<tbody>
<tr>
<td>L2a</td>
<td>3.3s</td>
<td>0.1s</td>
<td>0.1s</td>
<td>0.2s</td>
<td>1.0s</td>
<td>0.1s</td>
<td>1.8s</td>
</tr>
<tr>
<td>L2b</td>
<td>7.7s</td>
<td>1.3s</td>
<td>0.1s</td>
<td>0.3s</td>
<td>1.8s</td>
<td>1.9s</td>
<td>2.3s</td>
</tr>
<tr>
<td>L2c</td>
<td>3.9s</td>
<td>0.2s</td>
<td>0.2s</td>
<td>0.3s</td>
<td>0.7s</td>
<td>-</td>
<td>2.5s</td>
</tr>
<tr>
<td>L2d</td>
<td>2.8s</td>
<td>0.2s</td>
<td>0.0s</td>
<td>0.2s</td>
<td>0.4s</td>
<td>0.1s</td>
<td>1.8s</td>
</tr>
<tr>
<td>L2e</td>
<td>5.2s</td>
<td>0.3s</td>
<td>0.0s</td>
<td>0.3s</td>
<td>1.9s</td>
<td>0.5s</td>
<td>2.2s</td>
</tr>
<tr>
<td>L1a</td>
<td>5.7s</td>
<td>0.5s</td>
<td>0.1s</td>
<td>0.3s</td>
<td>1.8s</td>
<td>0.7s</td>
<td>2.2s</td>
</tr>
<tr>
<td>L1b</td>
<td>3.2s</td>
<td>0.1s</td>
<td>0.1s</td>
<td>0.2s</td>
<td>0.4s</td>
<td>0.1s</td>
<td>2.2s</td>
</tr>
<tr>
<td>L1c</td>
<td>5.8s</td>
<td>0.3s</td>
<td>0.2s</td>
<td>0.5s</td>
<td>1.5s</td>
<td>0.3s</td>
<td>3.1s</td>
</tr>
<tr>
<td>L1d</td>
<td>3.4s</td>
<td>0.7s</td>
<td>0.1s</td>
<td>0.3s</td>
<td>0.6s</td>
<td>-</td>
<td>1.8s</td>
</tr>
<tr>
<td>L1e</td>
<td>2.8s</td>
<td>0.3s</td>
<td>0.2s</td>
<td>0.2s</td>
<td>0.3s</td>
<td>0.1s</td>
<td>1.9s</td>
</tr>
</tbody>
</table>

L2 = native French, L1 = native English

average time per input word
### Time Spent on Activities

<table>
<thead>
<tr>
<th>User</th>
<th>total</th>
<th>initial</th>
<th>final</th>
<th>short</th>
<th>medium</th>
<th>big</th>
<th>keystroke</th>
</tr>
</thead>
<tbody>
<tr>
<td>L2a</td>
<td>3.3s</td>
<td>0.1s</td>
<td>0.1s</td>
<td>0.2s</td>
<td>1.0s</td>
<td>0.1s</td>
<td>1.8s</td>
</tr>
<tr>
<td>L2b</td>
<td>7.7s</td>
<td>1.3s</td>
<td>0.1s</td>
<td>0.3s</td>
<td>1.8s</td>
<td>1.9s</td>
<td>2.3s</td>
</tr>
<tr>
<td>L2c</td>
<td>3.9s</td>
<td>0.2s</td>
<td>0.2s</td>
<td>0.3s</td>
<td>0.7s</td>
<td></td>
<td>2.5s</td>
</tr>
<tr>
<td>L2d</td>
<td>2.8s</td>
<td>0.2s</td>
<td>0.0s</td>
<td>0.2s</td>
<td>0.4s</td>
<td>0.1s</td>
<td>1.8s</td>
</tr>
<tr>
<td>L2e</td>
<td>5.2s</td>
<td>0.3s</td>
<td>0.0s</td>
<td>0.3s</td>
<td>1.9s</td>
<td>0.5s</td>
<td>2.2s</td>
</tr>
<tr>
<td>L1a</td>
<td>5.7s</td>
<td>0.5s</td>
<td>0.1s</td>
<td>0.3s</td>
<td>1.8s</td>
<td>0.7s</td>
<td>2.2s</td>
</tr>
<tr>
<td>L1b</td>
<td>3.2s</td>
<td>0.1s</td>
<td>0.1s</td>
<td>0.2s</td>
<td>0.4s</td>
<td>0.1s</td>
<td>2.2s</td>
</tr>
<tr>
<td>L1c</td>
<td>5.8s</td>
<td>0.3s</td>
<td>0.2s</td>
<td>0.5s</td>
<td>1.5s</td>
<td>0.3s</td>
<td>3.1s</td>
</tr>
<tr>
<td>L1d</td>
<td>3.4s</td>
<td>0.7s</td>
<td>0.1s</td>
<td>0.3s</td>
<td>0.6s</td>
<td></td>
<td>1.8s</td>
</tr>
<tr>
<td>L1e</td>
<td>2.8s</td>
<td>0.3s</td>
<td>0.2s</td>
<td>0.2s</td>
<td>0.3s</td>
<td>0.1s</td>
<td>1.9s</td>
</tr>
</tbody>
</table>

L2 = native French, L1 = native English
average time per input word
Pauses Reconsidered

• Our classification of pauses is arbitrary (2-6sec, 6-60sec, >60sec)

• Extreme view: all you see is pauses
  – keystrokes take no observable time
  – all you see is pauses between action points
Pauses Reconsidered

• Our classification of pauses is arbitrary (2-6sec, 6-60sec, >60sec)

• Extreme view: all you see is pauses
  – keystrokes take no observable time
  – all you see is pauses between action points

• Visualizing range of pauses:
  time \( t \) spent in pauses \( p \in P \) up to a certain length \( l \)

\[
sum(t) = \frac{1}{Z} \sum_{p \in P, l(p) \leq t} l(p)
\]
Results

![Graph showing results]

The graph above illustrates the relationship between the length of pauses (in seconds) and the average time taken for translations across different scenarios labeled L1a to L2e.
Related Work: Tools used by Translators

- Translators often use standard text editors and additional tools
- Bilingual dictionary
- Spell checker, grammar checker
- Monolingual concordancer
- Terminology database
- Web search to establish and verify meaning of terms
Translation Memory

• Source:

This feature is available for free in the QX 3400.

• Fuzzy match in translation memory:

This feature is available for free in the QX 3200.
Diese Funktion ist kostenlos im Modell QX 3200 verfügbar.

• Translator inspects the fuzzy match and uses it in her translation.
# Bilingual Concordancer

<table>
<thead>
<tr>
<th>Examples</th>
<th>Windkraft (noun, feminine) (also: Windenergie)</th>
<th>wind power (noun)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Zum Vergleich: Windkraft schafft fast sieben Mal mehr.</td>
<td></td>
<td></td>
</tr>
<tr>
<td>By way of comparison, wind power generates almost seven times as much.</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Introducing Windcube, a new generation of wind Lidar for wind power.</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Windkraft ist eine etablierte, wettbewerbsfähige Technologie mit hoher Zuverlässigkeit.</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Wind power is an established, competitive technology with high reliability.</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Examples</th>
<th>Windkraft (noun, feminine) (also: Windenergie)</th>
<th>wind energy (noun)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Je mehr aber klimapolitische Sonntagsreden von der Politik auch in Taten umgesetzt werden, desto höher steigt dieser Preis und desto wettbewerbsfähiger werden saubere Energien wie die Windkraft.</td>
<td></td>
<td></td>
</tr>
<tr>
<td>But as the focus of the climate change issue shifts increasingly from policy to action, this price will increase and cleaner energy sources like wind will become more competitive.</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Nur wenige befürchten hingegen, dass dies auch bei erneuerbaren Energieträgern wie Biomasse oder Windkraft der Fall sein wird.</td>
<td></td>
<td></td>
</tr>
<tr>
<td>However, only a few fear that this will also be the case with renewable energy sources such as biomass or wind energy.</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

show translations in context (www.linguee.com)
Our Types of Assistance

- Sentence completion
  - tool suggests how to complete the translation
  - one phrase at a time
Our Types of Assistance

• Sentence completion
  – tool suggests how to complete the translation
  – one phrase at a time

• Translation options
  – most likely translations for each word and phrase
  – ordered and color-highlighted by probability
Our Types of Assistance

• Sentence completion
  – tool suggests how to complete the translation
  – one phrase at a time

• Translation options
  – most likely translations for each word and phrase
  – ordered and color-highlighted by probability

• Postediting machine translation
  – start with machine translation output
  – user edits, tool shows changes
Technical Notes

- Online at http://www.caitra.org/
- User uploads source text, translates one sentence at a time
- Implementation
  - AJAX Web 2.0 using Ruby on Rails, mySQL
  - Back end: Moses machine translation system
Predicting Sentence Completion

- Tool makes a suggestion how to continue (in red)
Predicting Sentence Completion

- Tool makes a suggestion how to continue (in red)
- User can accept it (by pressing TAB), or type in her own translation
Predicting Sentence Completion

- Tool makes a suggestion how to continue (in red)
- User can accept it (by pressing TAB), or type in her own translation
- Same idea as TransType, with minor modifications
  - show only short text chunks, not full sentence completion
  - show only one suggestion, not alternatives
How does it work?

- Uses search graph of SMT decoding
How does it work?

- Uses search graph of SMT decoding

- Matches partial user translation against search graph, by optimizing
  1. minimal string edit distance between path in graph and user translation
  2. best full path probability, including best completion to end
How does it work?

- Uses search graph of SMT decoding

- Matches partial user translation against search graph, by optimizing
  1. minimal string edit distance between path in graph and user translation
  2. best full path probability, including best completion to end

- Technical notes
  - search graph is pre-computed and stored in database
  - matching is done server-side, typically takes less than 1 second
  - completion path is returned to client (web browser)
Translation Options

- For each word and phrases: suggested translations
- Ranked (and color-highlighted) by probability
- User may click on suggestion → appended to text box
Translation Options - How does it work?

- Uses phrase translation table of SMT system
Translation Options - How does it work?

- Uses phrase translation table of SMT system

- Translation score: future cost estimate
  - conditional probabilities $\phi(\bar{e}|\bar{f}), \phi(\bar{f}|\bar{e})$
  - lexical probabilities $\text{lex}(\bar{e}|\bar{f}), \text{lex}(\bar{f}|\bar{e})$
  - word count feature
  - language model estimate
Translation Options - How does it work?

- Uses phrase translation table of SMT system

- Translation score: future cost estimate
  - conditional probabilities $\phi(\bar{e}|\bar{f})$, $\phi(\bar{f}|\bar{e})$
  - lexical probabilities $\text{lex}(\bar{e}|\bar{f})$, $\text{lex}(\bar{f}|\bar{e})$
  - word count feature
  - language model estimate

- Ranking of shorter vs. longer phrases by including outside future cost estimate
Sentence 2 of 20

[1] Spitzen von Hamburger CDU und Grünen öffnen Weg zu Koalitionsverhandlungen

[1] Leaders of the Hamburger CDU and Greens open path to coalition negotiations.
[2] The first black-green alliance on the state level is approaching: The leaders of CDU and Greens in Hamburg consider their differences to be surmountable. [3] In a preliminary phase, they discuss the start of coalition negotiations in the party caucuses. [4] Hamburg - They spent six hours discussing with each other. [5] Then the CDU-leader Michael Freytag and Green party leader Anja Hajduk announced that the division between the parties is bridgable.

Das erste schwarze-grüne Bündnis auf Landesebene rückt näher: Die Spitzen von CDU und Grünen in Hamburg halten ihre Differenzen für überwindbar.

<table>
<thead>
<tr>
<th>das</th>
<th>erste</th>
<th>schwar</th>
<th>Grüne</th>
<th>Bündnis</th>
<th>auf Landesebene</th>
<th>(rückt</th>
<th>näher)</th>
</tr>
</thead>
<tbody>
<tr>
<td>the</td>
<td>first</td>
<td>black</td>
<td>green</td>
<td>alliance</td>
<td>in favour of</td>
<td>is</td>
<td>that</td>
</tr>
<tr>
<td>for the first</td>
<td>black</td>
<td>Green</td>
<td>Alliance</td>
<td>on</td>
<td>national</td>
<td>approaches</td>
<td>the people at the top</td>
</tr>
<tr>
<td>this</td>
<td>in black and white</td>
<td>Grüne</td>
<td>cooperation</td>
<td>in</td>
<td>Belarussian approaches</td>
<td>we are coming to</td>
<td>at the top</td>
</tr>
<tr>
<td>the the first of</td>
<td>the black</td>
<td>the Greens</td>
<td>NATO</td>
<td>seek to</td>
<td>the top</td>
<td>the the top</td>
<td></td>
</tr>
</tbody>
</table>
Postediting Machine Translation

- Textbox is initially filled with machine translation
- User edits translation
- String edit distance to machine translation is shown (blue background)
Evaluation

- Recall setup
  - 10 students, half native French, half native English
  - each student translated French-English news stories
  - about 40 sentences for each condition of assistance
Evaluation

• Recall setup
  – 10 students, half native French, half native English
  – each student translated French-English news stories
  – about 40 sentences for each condition of assistance

• Five different conditions
  – unassisted
  – prediction (sentence completion)
  – options
  – predictions and options
  – post-editing
Quality

- We want faster translators, but not worse

- Assessment of translation quality
  - show translations to bilingual judges, with source
  - judgment: fully correct? yes/no

  Indicate whether each user’s input represents a fully fluent and meaning-equivalent translation of the source. The source is shown with context, the actual sentence is bold.
Quality

- We want faster translators, but not worse

- Assessment of translation quality
  - show translations to bilingual judges, with source
  - judgment: fully correct? yes/no

  Indicate whether each user’s input represents a fully fluent and meaning-equivalent translation of the source. The source is shown with context, the actual sentence is bold.

- Average score: 50% correct — lower than expected
  - judges seemed to be too harsh
  - when given several translations, tendency to judge half as bad
## Example of Quality Judgments

<table>
<thead>
<tr>
<th>Src.</th>
<th>Sans se démonter, il s’est montré concis et précis.</th>
</tr>
</thead>
<tbody>
<tr>
<td>MT</td>
<td>Without dismantle, it has been concise and accurate.</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Score</th>
<th>Translation</th>
<th>Notes</th>
</tr>
</thead>
<tbody>
<tr>
<td>1/3</td>
<td>Without fail, he has been concise and accurate.</td>
<td>(Prediction+Options, L1a)</td>
</tr>
<tr>
<td>4/0</td>
<td>Without getting flustered, he showed himself to be concise and precise.</td>
<td>(Unassisted, L1b)</td>
</tr>
<tr>
<td>4/0</td>
<td>Without falling apart, he has shown himself to be concise and accurate.</td>
<td>(Postedit, L1c)</td>
</tr>
<tr>
<td>1/3</td>
<td>Unswayable, he has shown himself to be concise and to the point.</td>
<td>(Options, L1d)</td>
</tr>
<tr>
<td>0/4</td>
<td>Without showing off, he showed himself to be concise and precise.</td>
<td>(Prediction, L1e)</td>
</tr>
<tr>
<td>1/3</td>
<td>Without dismantling himself, he presented himself consistent and precise.</td>
<td>(Prediction+Options, L2a)</td>
</tr>
<tr>
<td>2/2</td>
<td>He showed himself concise and precise.</td>
<td>(Unassisted, L2b)</td>
</tr>
<tr>
<td>3/1</td>
<td>Nothing daunted, he has been concise and accurate.</td>
<td>(Postedit, L2c)</td>
</tr>
<tr>
<td>3/1</td>
<td>Without losing face, he remained focused and specific.</td>
<td>(Options, L2d)</td>
</tr>
<tr>
<td>3/1</td>
<td>Without becoming flustered, he showed himself concise and precise.</td>
<td>(Prediction, L2e)</td>
</tr>
</tbody>
</table>
# Faster and Better

<table>
<thead>
<tr>
<th>Assistance</th>
<th>Speed</th>
<th>Quality</th>
</tr>
</thead>
<tbody>
<tr>
<td>Unassisted</td>
<td>4.4s/word</td>
<td>47% correct</td>
</tr>
<tr>
<td>Postedit</td>
<td>2.7s (-1.7s)</td>
<td>55% (+8%)</td>
</tr>
<tr>
<td>Options</td>
<td>3.7s (-0.7s)</td>
<td>51% (+4%)</td>
</tr>
<tr>
<td>Prediction</td>
<td>3.2s (-1.2s)</td>
<td>54% (+7%)</td>
</tr>
<tr>
<td>Prediction+Options</td>
<td>3.3s (-1.1s)</td>
<td>53% (+6%)</td>
</tr>
</tbody>
</table>
## Faster and Better, Mostly

<table>
<thead>
<tr>
<th>User</th>
<th>Unassisted</th>
<th>Postedit</th>
<th>Options</th>
<th>Prediction</th>
<th>Prediction+Options</th>
</tr>
</thead>
<tbody>
<tr>
<td>L2a</td>
<td>3.3s/word</td>
<td>1.2s</td>
<td>2.3s</td>
<td>1.1s</td>
<td>2.4s</td>
</tr>
<tr>
<td></td>
<td>23% correct</td>
<td>-2.2s</td>
<td>-1.0s</td>
<td>-2.2s</td>
<td>-0.9s</td>
</tr>
<tr>
<td></td>
<td></td>
<td>39%</td>
<td>45%</td>
<td>30%</td>
<td>44%</td>
</tr>
<tr>
<td></td>
<td></td>
<td>+16%</td>
<td>+22%</td>
<td>+7%</td>
<td>+21%</td>
</tr>
<tr>
<td>L2b</td>
<td>7.7s/word</td>
<td>4.5s</td>
<td>4.5s</td>
<td>2.7s</td>
<td>4.8s</td>
</tr>
<tr>
<td></td>
<td>35% correct</td>
<td>-3.2s</td>
<td>-3.3s</td>
<td>-5.1s</td>
<td>-3.0s</td>
</tr>
<tr>
<td></td>
<td></td>
<td>48%</td>
<td>55%</td>
<td>61%</td>
<td>41%</td>
</tr>
<tr>
<td></td>
<td></td>
<td>+13%</td>
<td>+20%</td>
<td>+26%</td>
<td>+6%</td>
</tr>
<tr>
<td>L2c</td>
<td>3.9s/word</td>
<td>1.9s</td>
<td>3.8s</td>
<td>3.1s</td>
<td>2.5s</td>
</tr>
<tr>
<td></td>
<td>50% correct</td>
<td>-2.0s</td>
<td>-0.1s</td>
<td>-0.8s</td>
<td>-1.4s</td>
</tr>
<tr>
<td></td>
<td></td>
<td>61%</td>
<td>54%</td>
<td>64%</td>
<td>61%</td>
</tr>
<tr>
<td></td>
<td></td>
<td>+11%</td>
<td>+4%</td>
<td>+14%</td>
<td>+11%</td>
</tr>
<tr>
<td>L2d</td>
<td>2.8s/word</td>
<td>2.0s</td>
<td>2.9s</td>
<td>2.4s</td>
<td>1.8s</td>
</tr>
<tr>
<td></td>
<td>38% correct</td>
<td>-0.7s</td>
<td>(+0.1s)</td>
<td>(-0.4s)</td>
<td>(1%)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>46%</td>
<td>59%</td>
<td>37%</td>
<td>45%</td>
</tr>
<tr>
<td></td>
<td></td>
<td>+8%</td>
<td>(+21%)</td>
<td>(-1%)</td>
<td>+7%</td>
</tr>
<tr>
<td>L2e</td>
<td>5.2s/word</td>
<td>3.9s</td>
<td>4.9s</td>
<td>3.5s</td>
<td>4.6s</td>
</tr>
<tr>
<td></td>
<td>58% correct</td>
<td>-1.3s</td>
<td>(-0.2s)</td>
<td>-1.7s</td>
<td>(-0.5s)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>64%</td>
<td>56%</td>
<td>62%</td>
<td>56%</td>
</tr>
<tr>
<td></td>
<td></td>
<td>+6%</td>
<td>(-2%)</td>
<td>+4%</td>
<td>(-2%)</td>
</tr>
<tr>
<td>L1a</td>
<td>5.7s/word</td>
<td>1.8s</td>
<td>2.5s</td>
<td>2.7s</td>
<td>2.8s</td>
</tr>
<tr>
<td></td>
<td>16% correct</td>
<td>-3.9s</td>
<td>-3.2s</td>
<td>-3.0s</td>
<td>-2.9s</td>
</tr>
<tr>
<td></td>
<td></td>
<td>50%</td>
<td>34%</td>
<td>40%</td>
<td>50%</td>
</tr>
<tr>
<td></td>
<td></td>
<td>+34%</td>
<td>+18%</td>
<td>+24%</td>
<td>+34%</td>
</tr>
<tr>
<td>L1b</td>
<td>3.2s/word</td>
<td>2.8s</td>
<td>3.5s</td>
<td>6.0s</td>
<td>4.6s</td>
</tr>
<tr>
<td></td>
<td>64% correct</td>
<td>(-0.4s)</td>
<td>+0.3s</td>
<td>+2.8s</td>
<td>+1.4s</td>
</tr>
<tr>
<td></td>
<td></td>
<td>56%</td>
<td>60%</td>
<td>61%</td>
<td>57%</td>
</tr>
<tr>
<td></td>
<td></td>
<td>(-8%)</td>
<td>-4%</td>
<td>-3%</td>
<td>-7%</td>
</tr>
<tr>
<td>L1c</td>
<td>5.8s/word</td>
<td>2.9s</td>
<td>4.6s</td>
<td>4.1s</td>
<td>2.7s</td>
</tr>
<tr>
<td></td>
<td>52% correct</td>
<td>-3.0s</td>
<td>(-1.2s)</td>
<td>-1.7s</td>
<td>-3.1s</td>
</tr>
<tr>
<td></td>
<td></td>
<td>53%</td>
<td>37%</td>
<td>59%</td>
<td>53%</td>
</tr>
<tr>
<td></td>
<td></td>
<td>+1%</td>
<td>(-15%)</td>
<td>+7%</td>
<td>+1%</td>
</tr>
<tr>
<td>L1d</td>
<td>3.4s/word</td>
<td>3.1s</td>
<td>4.3s</td>
<td>3.8s</td>
<td>3.7s</td>
</tr>
<tr>
<td></td>
<td>49% correct</td>
<td>(-0.3s)</td>
<td>(+0.9s)</td>
<td>(+0.4s)</td>
<td>(+0.3s)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>49%</td>
<td>51%</td>
<td>53%</td>
<td>58%</td>
</tr>
<tr>
<td></td>
<td></td>
<td>(-0%)</td>
<td>(+2%)</td>
<td>(-4%)</td>
<td>(+9%)</td>
</tr>
<tr>
<td>L1e</td>
<td>2.8s/word</td>
<td>2.6s</td>
<td>3.5s</td>
<td>2.8s</td>
<td>3.0s</td>
</tr>
<tr>
<td></td>
<td>68% correct</td>
<td>-0.2s</td>
<td>+0.7s</td>
<td>(-0.0s)</td>
<td>+0.2s</td>
</tr>
<tr>
<td></td>
<td></td>
<td>79%</td>
<td>59%</td>
<td>64%</td>
<td>66%</td>
</tr>
<tr>
<td></td>
<td></td>
<td>+11%</td>
<td>-9%</td>
<td>(-4%)</td>
<td>-2%</td>
</tr>
<tr>
<td>avg.</td>
<td>4.4s/word</td>
<td>2.7s</td>
<td>3.7s</td>
<td>3.2s</td>
<td>3.3s</td>
</tr>
<tr>
<td></td>
<td>47% correct</td>
<td>-1.7s</td>
<td>-0.7s</td>
<td>-1.2s</td>
<td>-1.1s</td>
</tr>
<tr>
<td></td>
<td></td>
<td>55%</td>
<td>51%</td>
<td>54%</td>
<td>53%</td>
</tr>
<tr>
<td></td>
<td></td>
<td>+8%</td>
<td>+4%</td>
<td>+7%</td>
<td>+6%</td>
</tr>
</tbody>
</table>
Slow Users 1: Faster and Better

- Unassisted
  - more than 5 seconds per input word
  - very bad (35%, 16%)

- With assistance
  - much faster and better
  - reaching roughly average performance
Slow Users 2: Only Faster

- Unassisted
  - more than 5 seconds per input word
  - average quality

- With assistance
  - faster and but not better
• Unassisted
  – fast: 3-4 seconds per input word
  – L2a is very bad (23%), L2c is average (50%)

• With assistance
  – faster and better
  – L2a closer to average (30-45%), L2c becomes very good (54-61%)
• Use the assistance sparingly or not at all, and see generally no gains

• The two best translators are in this group

• Postediting
  – mixed on quality (2 better, 1 worse, 1 same), but all faster
  – best translator (L1e, 68%) becomes much better (record 79%)
Further Analysis

- How does the assistance change translator behaviour?
- How do translators utilize assistance?
- How is the translation produced?
Keystroke Log

black: keystroke, purple: deletion, grey: cursor move
red: sentence completion accept
orange: click on translation option

Analysis: Segment into periods of activity: typing, tabbing, clicking, pauses

one second before and after a keystroke is part of typing interval
### Activities: Native French User L2b

<table>
<thead>
<tr>
<th>User: L2b</th>
<th>total</th>
<th>init-p</th>
<th>end-p</th>
<th>short-p</th>
<th>mid-p</th>
<th>big-p</th>
<th>key</th>
<th>click</th>
<th>tab</th>
</tr>
</thead>
<tbody>
<tr>
<td>Unassisted</td>
<td>7.7s</td>
<td>1.3s</td>
<td>0.1s</td>
<td>0.3s</td>
<td>1.8s</td>
<td>1.9s</td>
<td>2.3s</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>Postedit</td>
<td>4.5s</td>
<td>1.5s</td>
<td>0.4s</td>
<td>0.1s</td>
<td>1.0s</td>
<td>0.4s</td>
<td>1.1s</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>Options</td>
<td>4.5s</td>
<td>0.6s</td>
<td>0.1s</td>
<td>0.4s</td>
<td>0.9s</td>
<td>0.7s</td>
<td>1.5s</td>
<td>0.4s</td>
<td>-</td>
</tr>
<tr>
<td>Prediction</td>
<td>2.7s</td>
<td>0.3s</td>
<td>0.3s</td>
<td>0.2s</td>
<td>0.7s</td>
<td>0.1s</td>
<td>0.6s</td>
<td>-</td>
<td>0.4s</td>
</tr>
<tr>
<td>Prediction+Options</td>
<td>4.8s</td>
<td>0.6s</td>
<td>0.4s</td>
<td>0.4s</td>
<td>1.3s</td>
<td>0.5s</td>
<td>0.9s</td>
<td>0.5s</td>
<td>0.2s</td>
</tr>
</tbody>
</table>
## Activities: Native French User L2b

<table>
<thead>
<tr>
<th>User: L2b</th>
<th>total</th>
<th>init-p</th>
<th>end-p</th>
<th>short-p</th>
<th>mid-p</th>
<th>big-p</th>
<th>key</th>
<th>click</th>
<th>tab</th>
</tr>
</thead>
<tbody>
<tr>
<td>Unassisted</td>
<td>7.7s</td>
<td>1.3s</td>
<td>0.1s</td>
<td>0.3s</td>
<td>1.8s</td>
<td>1.9s</td>
<td>2.3s</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>Postedit</td>
<td>4.5s</td>
<td>1.5s</td>
<td>0.4s</td>
<td>0.1s</td>
<td>1.0s</td>
<td>0.4s</td>
<td>1.1s</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>Options</td>
<td>4.5s</td>
<td>0.6s</td>
<td>0.1s</td>
<td>0.4s</td>
<td>0.9s</td>
<td>0.7s</td>
<td>1.5s</td>
<td>0.4s</td>
<td>-</td>
</tr>
<tr>
<td>Prediction</td>
<td>2.7s</td>
<td>0.3s</td>
<td>0.3s</td>
<td>0.2s</td>
<td>0.7s</td>
<td>0.1s</td>
<td>0.6s</td>
<td>-</td>
<td>0.4s</td>
</tr>
<tr>
<td>Prediction+Options</td>
<td>4.8s</td>
<td>0.6s</td>
<td>0.4s</td>
<td>0.4s</td>
<td>1.3s</td>
<td>0.5s</td>
<td>0.9s</td>
<td>0.5s</td>
<td>0.2s</td>
</tr>
</tbody>
</table>

Slightly less time spent on typing.
## Activities: Native French User L2b

<table>
<thead>
<tr>
<th>User: L2b</th>
<th>total</th>
<th>init-p</th>
<th>end-p</th>
<th>short-p</th>
<th>mid-p</th>
<th>big-p</th>
<th>key</th>
<th>click</th>
<th>tab</th>
</tr>
</thead>
<tbody>
<tr>
<td>Unassisted</td>
<td>7.7s</td>
<td>1.3s</td>
<td>0.1s</td>
<td>0.3s</td>
<td>1.8s</td>
<td>1.9s</td>
<td>2.3s</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>Postedit</td>
<td>4.5s</td>
<td>1.5s</td>
<td>0.4s</td>
<td>0.1s</td>
<td>1.0s</td>
<td>0.4s</td>
<td>1.1s</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>Options</td>
<td>4.5s</td>
<td>0.6s</td>
<td>0.1s</td>
<td>0.4s</td>
<td>0.9s</td>
<td>0.7s</td>
<td>1.5s</td>
<td>0.4s</td>
<td>-</td>
</tr>
<tr>
<td>Prediction</td>
<td>2.7s</td>
<td>0.3s</td>
<td>0.3s</td>
<td>0.2s</td>
<td>0.7s</td>
<td>0.1s</td>
<td>0.6s</td>
<td>-</td>
<td>0.4s</td>
</tr>
<tr>
<td>Prediction+Options</td>
<td>4.8s</td>
<td>0.6s</td>
<td>0.4s</td>
<td>0.4s</td>
<td>1.3s</td>
<td>0.5s</td>
<td>0.9s</td>
<td>0.5s</td>
<td>0.2s</td>
</tr>
</tbody>
</table>

Less pausing

Slightly less time spent on typing
Activities: Native French User L2b

<table>
<thead>
<tr>
<th>User: L2b</th>
<th>total</th>
<th>init-p</th>
<th>end-p</th>
<th>short-p</th>
<th>mid-p</th>
<th>big-p</th>
<th>key</th>
<th>click</th>
<th>tab</th>
</tr>
</thead>
<tbody>
<tr>
<td>Unassisted</td>
<td>7.7s</td>
<td>1.3s</td>
<td>0.1s</td>
<td>0.3s</td>
<td>1.8s</td>
<td>1.9s</td>
<td>2.3s</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>Postedit</td>
<td>4.5s</td>
<td>1.5s</td>
<td>0.4s</td>
<td>0.1s</td>
<td>1.0s</td>
<td>0.4s</td>
<td>1.1s</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>Options</td>
<td>4.5s</td>
<td>0.6s</td>
<td>0.1s</td>
<td>0.4s</td>
<td>0.9s</td>
<td>0.7s</td>
<td>1.5s</td>
<td>0.4s</td>
<td>-</td>
</tr>
<tr>
<td>Prediction</td>
<td>2.7s</td>
<td>0.3s</td>
<td>0.3s</td>
<td>0.2s</td>
<td>0.7s</td>
<td>0.1s</td>
<td>0.6s</td>
<td>-</td>
<td>0.4s</td>
</tr>
<tr>
<td>Prediction+Options</td>
<td>4.8s</td>
<td>0.6s</td>
<td>0.4s</td>
<td>0.4s</td>
<td>1.3s</td>
<td>0.5s</td>
<td>0.9s</td>
<td>0.5s</td>
<td>0.2s</td>
</tr>
</tbody>
</table>

- Less pausing
- Especially less time in big pauses
- Slightly less time spent on typing
# Activities: Native English User L1e

<table>
<thead>
<tr>
<th>User: L1e</th>
<th>total</th>
<th>init-p</th>
<th>end-p</th>
<th>short-p</th>
<th>mid-p</th>
<th>big-p</th>
<th>key</th>
<th>click</th>
<th>tab</th>
</tr>
</thead>
<tbody>
<tr>
<td>Unassisted</td>
<td>2.8s</td>
<td>0.3s</td>
<td>0.2s</td>
<td>0.2s</td>
<td>0.3s</td>
<td>0.1s</td>
<td>1.9s</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>Postedit</td>
<td>2.6s</td>
<td>0.4s</td>
<td>0.3s</td>
<td>0.2s</td>
<td>1.0s</td>
<td>0.1s</td>
<td>0.7s</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>Options</td>
<td>3.5s</td>
<td>0.1s</td>
<td>0.3s</td>
<td>0.4s</td>
<td>0.6s</td>
<td>0.2s</td>
<td>1.7s</td>
<td>0.1s</td>
<td>-</td>
</tr>
<tr>
<td>Prediction</td>
<td>2.8s</td>
<td>0.1s</td>
<td>0.3s</td>
<td>0.3s</td>
<td>0.3s</td>
<td>-</td>
<td>1.4s</td>
<td>-</td>
<td>0.3s</td>
</tr>
<tr>
<td>Prediction+Options</td>
<td>3.0s</td>
<td>0.1s</td>
<td>0.3s</td>
<td>0.2s</td>
<td>0.5s</td>
<td>-</td>
<td>1.9s</td>
<td>-</td>
<td>-</td>
</tr>
</tbody>
</table>
## Activities: Native English User L1e

<table>
<thead>
<tr>
<th>User: L1e</th>
<th>total</th>
<th>init-p</th>
<th>end-p</th>
<th>short-p</th>
<th>mid-p</th>
<th>big-p</th>
<th>key</th>
<th>click</th>
<th>tab</th>
</tr>
</thead>
<tbody>
<tr>
<td>Unassisted</td>
<td>2.8s</td>
<td>0.3s</td>
<td>0.2s</td>
<td>0.2s</td>
<td>0.3s</td>
<td>0.1s</td>
<td>1.9s</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>Postedit</td>
<td>2.6s</td>
<td>0.4s</td>
<td>0.3s</td>
<td>0.2s</td>
<td>1.0s</td>
<td>0.1s</td>
<td>0.7s</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>Options</td>
<td>3.5s</td>
<td>0.1s</td>
<td>0.3s</td>
<td>0.4s</td>
<td>0.6s</td>
<td>0.2s</td>
<td>1.7s</td>
<td>0.1s</td>
<td>-</td>
</tr>
<tr>
<td>Prediction</td>
<td>2.8s</td>
<td>0.1s</td>
<td>0.3s</td>
<td>0.3s</td>
<td>0.3s</td>
<td>-</td>
<td>1.4s</td>
<td>-</td>
<td>0.3s</td>
</tr>
<tr>
<td>Prediction+Options</td>
<td>3.0s</td>
<td>0.1s</td>
<td>0.3s</td>
<td>0.2s</td>
<td>0.5s</td>
<td>-</td>
<td>1.9s</td>
<td>-</td>
<td>-</td>
</tr>
</tbody>
</table>

Little time spent on assistance
### Activities: Native English User L1e

<table>
<thead>
<tr>
<th>User: L1e</th>
<th>total</th>
<th>init-p</th>
<th>end-p</th>
<th>short-p</th>
<th>mid-p</th>
<th>big-p</th>
<th>key</th>
<th>click</th>
<th>tab</th>
</tr>
</thead>
<tbody>
<tr>
<td>Unassisted</td>
<td>2.8s</td>
<td>0.3s</td>
<td>0.2s</td>
<td>0.2s</td>
<td>0.3s</td>
<td>0.1s</td>
<td>1.9s</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>Postedit</td>
<td>2.6s</td>
<td>0.4s</td>
<td>0.3s</td>
<td>0.2s</td>
<td>1.0s</td>
<td>0.1s</td>
<td>0.7s</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>Options</td>
<td>3.5s</td>
<td>0.1s</td>
<td>0.3s</td>
<td>0.4s</td>
<td>0.6s</td>
<td>0.2s</td>
<td>1.7s</td>
<td>0.1s</td>
<td>-</td>
</tr>
<tr>
<td>Prediction</td>
<td>2.8s</td>
<td>0.1s</td>
<td>0.3s</td>
<td>0.3s</td>
<td>0.3s</td>
<td>-</td>
<td>1.4s</td>
<td>-</td>
<td>0.3s</td>
</tr>
<tr>
<td>Prediction+Options</td>
<td>3.0s</td>
<td>0.1s</td>
<td>0.3s</td>
<td>0.2s</td>
<td>0.5s</td>
<td>-</td>
<td>1.9s</td>
<td>-</td>
<td>-</td>
</tr>
</tbody>
</table>

- Does not use both assistances, little overall change
- Little time spent on assistance
### Activities: Native English User L1e

<table>
<thead>
<tr>
<th>User: L1e</th>
<th>total</th>
<th>init-p</th>
<th>end-p</th>
<th>short-p</th>
<th>mid-p</th>
<th>big-p</th>
<th>key</th>
<th>click</th>
<th>tab</th>
</tr>
</thead>
<tbody>
<tr>
<td>Unassisted</td>
<td>2.8s</td>
<td>0.3s</td>
<td>0.2s</td>
<td>0.2s</td>
<td>0.3s</td>
<td>0.1s</td>
<td>1.9s</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>Postedit</td>
<td>2.6s</td>
<td>0.4s</td>
<td>0.3s</td>
<td>0.2s</td>
<td>1.0s</td>
<td>0.1s</td>
<td>0.7s</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>Options</td>
<td>3.5s</td>
<td>0.1s</td>
<td>0.3s</td>
<td>0.4s</td>
<td>0.6s</td>
<td>0.2s</td>
<td>1.7s</td>
<td>0.1s</td>
<td>-</td>
</tr>
<tr>
<td>Prediction</td>
<td>2.8s</td>
<td>0.1s</td>
<td>0.3s</td>
<td>0.3s</td>
<td>0.3s</td>
<td>-</td>
<td>1.4s</td>
<td>-</td>
<td>0.3s</td>
</tr>
<tr>
<td>Prediction+Options</td>
<td>3.0s</td>
<td>0.1s</td>
<td>0.3s</td>
<td>0.2s</td>
<td>0.5s</td>
<td>-</td>
<td>1.9s</td>
<td>-</td>
<td>-</td>
</tr>
</tbody>
</table>

- **Does not use both assistances, little overall change**
- **Postediting:**
  - less typing (-1.2s)
  - more medium pauses (+0.7s)
- **Little time spent on assistance**
Origin of Characters: Native French L2b

<table>
<thead>
<tr>
<th>User: L2b</th>
<th>key</th>
<th>click</th>
<th>tab</th>
<th>mt</th>
</tr>
</thead>
<tbody>
<tr>
<td>Postedit</td>
<td>18%</td>
<td>-</td>
<td>-</td>
<td>81%</td>
</tr>
<tr>
<td>Options</td>
<td>59%</td>
<td>40%</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>Prediction</td>
<td>14%</td>
<td>-</td>
<td>85%</td>
<td>-</td>
</tr>
<tr>
<td>Prediction + Options</td>
<td>21%</td>
<td>44%</td>
<td>33%</td>
<td>-</td>
</tr>
</tbody>
</table>
Origin of Characters: Native French L2b

<table>
<thead>
<tr>
<th>User: L2b</th>
<th>key</th>
<th>click</th>
<th>tab</th>
<th>mt</th>
</tr>
</thead>
<tbody>
<tr>
<td>Postedit</td>
<td>18%</td>
<td>-</td>
<td>-</td>
<td>81%</td>
</tr>
<tr>
<td>Options</td>
<td>59%</td>
<td>40%</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>Prediction</td>
<td>14%</td>
<td>-</td>
<td>85%</td>
<td>-</td>
</tr>
<tr>
<td>Prediction+Options</td>
<td>21%</td>
<td>44%</td>
<td>33%</td>
<td>-</td>
</tr>
</tbody>
</table>

Translation comes to large degree from assistance
## Origin of Characters: Native English L1e

<table>
<thead>
<tr>
<th>User: L1e</th>
<th>key</th>
<th>click</th>
<th>tab</th>
<th>mt</th>
</tr>
</thead>
<tbody>
<tr>
<td>Postedit</td>
<td>20%</td>
<td>-</td>
<td>-</td>
<td>79%</td>
</tr>
<tr>
<td>Options</td>
<td>77%</td>
<td>22%</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>Prediction</td>
<td>61%</td>
<td>-</td>
<td>38%</td>
<td>-</td>
</tr>
<tr>
<td>Prediction+Options</td>
<td>100%</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
</tbody>
</table>
### Origin of Characters: Native English L1e

<table>
<thead>
<tr>
<th>User: L1e</th>
<th>key</th>
<th>click</th>
<th>tab</th>
<th>mt</th>
</tr>
</thead>
<tbody>
<tr>
<td>Postedit</td>
<td>20%</td>
<td>-</td>
<td>-</td>
<td>79%</td>
</tr>
<tr>
<td>Options</td>
<td>77%</td>
<td>22%</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>Prediction</td>
<td>61%</td>
<td>-</td>
<td>38%</td>
<td>-</td>
</tr>
<tr>
<td>Prediction+Options</td>
<td>100%</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
</tbody>
</table>

Although hardly any time spent on assistance, fair amount of characters produced by it.
Pauses: French-Native User L2b

![Graph showing average translations time (sec/word) vs. length of pauses (sec). The graph compares different conditions: Unassisted, Postedit, Options, Prediction, and Prediction+Options. The data shows a trend where longer pauses lead to increased time per word for all conditions, with Unassisted taking the longest and Prediction+Options the shortest, especially at longer pause lengths.](image-url)
Pauses: English-Native User L1e

The diagram shows the average translation time (sec/word) as a function of the length of pauses (sec). The lines represent the following conditions:

- Unassisted
- Postedit
- Options
- Prediction
- Prediction+Options

The x-axis represents the length of pauses in seconds, ranging from 0.1 to 100 seconds. The y-axis represents the average translation time per word, ranging from 0 to 8 seconds.
Learning Curve

users become better over time with assistance
User Feedback

- Q: In which of the five conditions did you think you were most accurate?
  - predictions+options: 5 users
  - options: 2 users
  - prediction: 1 user
  - postediting: 1 user
User Feedback

• Q: In which of the five conditions did you think you were most accurate?
  – predictions+options: 5 users
  – options: 2 users
  – prediction: 1 user
  – postediting: 1 user

• Q: Rank the different types of assistance on a scale from 1 to 5, where 1 indicates not at all and 5 indicates very helpful.
  – prediction+options: 4.6
  – prediction: 3.9
  – options: 3.7
  – postediting: 2.9
User Feedback

• Q: In which of the five conditions did you think you were most accurate?
  – predictions+options: 5 users
  – options: 2 users
  – prediction: 1 user
  – postediting: 1 user

• Q: Rank the different types of assistance on a scale from 1 to 5, where 1 indicates not at all and 5 indicates very helpful.
  – prediction+options: 4.6
  – prediction: 3.9
  – options: 3.7
  – postediting: 2.9
User Feedback

- Q: In which of the five conditions did you think you were most accurate?
  - predictions+options: 5 users
  - options: 2 users
  - prediction: 1 user
  - postediting: 1 user

- Q: Rank the different types of assistance on a scale from 1 to 5, where 1 indicates not at all and 5 indicates very helpful.
  - prediction+options: 4.6
  - prediction: 3.9
  - options: 3.7
  - postediting: 2.9

- Note: does not match empirical results
Summary

- Assistance made translators faster
  - average speed improvement from 4.4s/word to 2.7-3.7s/word
  - reduction of big pauses
  - reduction of typing effort in post-editing
Summary

• Assistance made translators faster
  – average speed improvement from 4.4s/word to 2.7-3.7s/word
  – reduction of big pauses
  – reduction of typing effort in post-editing

• Assistance made translators better
  – average judgment increased from 47% to 51-55% with help
  – even good translators get better with postediting
Summary

- Assistance made translators faster
  - average speed improvement from 4.4s/word to 2.7-3.7s/word
  - reduction of big pauses
  - reduction of typing effort in post-editing

- Assistance made translators better
  - average judgment increased from 47% to 51-55% with help
  - even good translators get better with postediting

- Some good translators ignored the assistance
Summary

• Assistance made translators faster
  – average speed improvement from 4.4s/word to 2.7-3.7s/word
  – reduction of big pauses
  – reduction of typing effort in post-editing

• Assistance made translators better
  – average judgment increased from 47% to 51-55% with help
  – even good translators get better with postediting

• Some good translators ignored the assistance

• Fastest and (barely) best with postediting, but did not like it
Outlook: More analysis

• What do translators think about when they are pausing?
Outlook: More analysis

- What do translators think about when they are pausing?

- What are the hard problems?
  - unknown words
  - words without direct translation
  - syntactic re-arrangement
Outlook: More analysis

• What do translators think about when they are pausing?

• What are the hard problems?
  – unknown words
  – words without direct translation
  – syntactic re-arrangement

• What do translators change in post-editing?
Outlook: More experiments

- Different types of users
  - experienced professional translators
  - volunteer / amateur
  - no/little knowledge of source language
Outlook: More experiments

- Different types of users
  - experienced professional translators
  - volunteer / amateur
  - no/little knowledge of source language

- Different types of language pairs
  - target-side morphology a problem
  - large-scale reordering maybe a problem
Outlook: More experiments

- Different types of users
  - experienced professional translators
  - volunteer / amateur
  - no/little knowledge of source language

- Different types of language pairs
  - target-side morphology a problem
  - large-scale reordering maybe a problem

- Different types of translation tasks
  - content unfamiliar to translator
  - test material very similar to previously translated text
Outlook: More experiments

- Different types of users
  - experienced professional translators
  - volunteer / amateur
  - no/little knowledge of source language

- Different types of language pairs
  - target-side morphology a problem
  - large-scale reordering maybe a problem

- Different types of translation tasks
  - content unfamiliar to translator
  - test material very similar to previously translated text
Outlook: Better Assistance

- Confidence estimation for post-editing
  - we have training data: pairs of MT output, corrected sentence
  → supervised binary classification problem (correct/false)
Outlook: Better Assistance

• Confidence estimation for post-editing
  – we have training data: pairs of MT output, corrected sentence
  → supervised binary classification problem (correct/false)

• Visualizing word alignment (not only for post-editing)
Outlook: Better Assistance

- Confidence estimation for post-editing
  - we have training data: pairs of MT output, corrected sentence
  → supervised binary classification problem (correct/false)

- Visualizing word alignment (not only for post-editing)

- Better prediction model
  - different metrics to find best matching prefix
    * BLEU, TER, ...
    * word substitution scored by letter string edit distance
  - prediction as a machine learning multi-class classification problem
  - prediction based on parse forest from tree-based models
Outlook: Additional Types of Assistance

- Bilingual concordancer
  - show translations of words and phrases in context
Outlook: Additional Types of Assistance

- Bilingual concordancer
  - show translations of words and phrases in context

- Monolingual concordancer / language model
  - show which expressions are more or less common
  - language model to highlight unusual transitions in output
Outlook: Additional Types of Assistance

• Bilingual concordancer
  – show translations of words and phrases in context

• Monolingual concordancer / language model
  – show which expressions are more or less common
  – language model to highlight unusual transitions in output

• Translation memory
  – finding most similar sentence in parallel corpus
  – show translation, highlight difference, fill in difference with MT
Try it at home!

http://www.caitra.org/

questions?
Pauses: Unassisted

![Graph showing pauses in translation time against pause length for different samples L1a to L2e.]

Philipp Koehn

Computer Aided Translation

29 September 2009
Pauses: Options

![Graph showing average translation time over length of pauses for different options](graph.png)
Pauses: Prediction of sentence completion

The diagram shows the relationship between the average translations time (sec/word) and the length of pauses (sec). The lines represent different groups labeled L1a, L1b, L1c, L1d, L1e, L2a, L2b, L2c, L2d, and L2e. The time generally increases with the length of pauses, indicating that longer pauses lead to longer translation times.
Pauses: Postediting

![Graph showing average translation times vs. length of pauses (sec).]